Inferring Ground Truth From Crowdsourced Data Under Local Attribute Differential Privacy

Di Wang[1,2] and Jinhui Xu[1]

[1] Department of Computer Science and Engineering
State University of New York at Buffalo, Buffalo, NY 14260, USA
[2] King Abdullah University of Science and Technology, Thuwal, Saudi Arabia
{dwang45,jinhui}@buffalo.edu

Abstract. Nowadays, crowdsourcing gains an increasing popularity as it can be adopted to solve many challenging question answering tasks that are easy for humans but difficult for computers. Due to the variety in the quality of users, it is important to infer not only the underlying ground truth of these tasks but also the users ability from the answers given by users. This problem is called Ground Truth Inference and has been studied for many years. However, since the answers collected from the users may contain sensitive information, ground truth inference raises serious privacy concern. Due to this reason, the problem of ground truth inference under local differential privacy (LDP) model has been recently studied. However, this problem is still not well understood and even some basic questions have not been solved yet. First, it is still unknown what is the average error of the private estimators to the underlying ground truth. Secondly, we do not known whether we can infer the ability of each user under LDP model and what is the estimation error w.r.t the underlying users ability. Finally, previous work only show that their methods have better performance than the private major voting algorithm through experiments. However, there is still no theoretically result which shows this priority formally or mathematically. In this paper, we partially solve these problems by studying the ground truth inference problem under local attribute differential privacy (LADP) model, which is a relaxation of LDP model, and propose a new algorithm called private Dawid-Skene method, which is motivated by the classical Dawid-Skene method. Specifically, we first provide the estimation errors for both ability of users and the ground truth under some assumptions of the problem if the algorithm start with some appropriate initial vector. Moreover, we propose an explicit instance and show that the estimation error of the ground truth achieved by the private major voting algorithm is always greater than the error achieved by our method.

Keywords: Differential Privacy · Crowdsourcing · Dawid-Skene Estimator

1 Introduction

Nowadays, crowdsourcing gains an increasing popularity as it can be adopted to solve many challenging question answering tasks that are easy for humans but difficult for the computer, and it has many real-world machine learning or data mining applications. For example, patients who are taking new drugs can answer the question on whether a specific drug has a certain side-effect [19]. Also there are many commercial web service for crowdsourcing such as Amazon Mechanical Turk (AMT). In these and many more applications, crowds of users can contribute their efforts to answer questions of interest, which largely reduces the financial cost and benefits various application domains.

Due to the variety in the quality of users, the information quality of the answers given by the users varies significantly. Some users may have sufficient domain knowledge and can provide accurate answers while others may submit biased or wrong answers. This diversity of users motivates a basic and important problem in crowdsourcing: how the do the server get the accurate answers (or ground truth) via these noisy answers while also could infer the underlying ability of each user. This problem is called **Ground Truth Inference**[3] [25] and there is a large amount of work study this problem in both Machine Learning [24], Data Mining [26] and Theoretical Computer Science [3, 12, 15] communities.

However, in the problem of ground truth inference, collecting individual users answers may cause the privacy issue on the users. For example, individual users can report the relevance between a search query and a webpage, but their answers may leak their personal preference. Patients' reactions to drugs are valuable for physicians to discover drugs' side-effect, but these also contain sensitive information. Moreover, recently it has been reported that AMT platform was leveraged by politicians to access a large pool of Facebook profiles and collects ten of thousands of individuals demographic data [22].

As a strong mathematical scheme for privacy-preserving, Differential Privacy [4] recently has been used in a many applications on machine learning and data mining and is now becoming a standard in data analysis. Two main user models have emerged for differential privacy: the central model and the local one. In the central model, data are managed by a trusted central entity which is responsible for collecting them and for deciding which differentially private data analysis to perform and to release. A classical use case for this model is the one for collecting census data [8]. In the local model, each individual manages his/her proper data and discloses them to a server through some differentially private mechanisms. The server collects the (now private) data of each individual and combines them into a resulting data analysis. A classical application of this model is the one aiming at collecting statistics from user devices like in the case of Google's Chrome browser [6], and Apple's iOS-10 [23]. Thus, compared with the central model, we can see that the ground truth inference is more suitable for the local differential privacy (LDP) model.

Ground truth inference in LDP model has been first studied by [16] and was later extended by [22] to the sparse crowdsourcing data case. Although their methods are effective with tolerable accuracy loss practically, there are still some basic theoretical open problems which have not been studied or solved. First, it is still unknown what is the average error of the private estimators with respect to the underlying ground truth. Secondly, while all the previous work focus on the quality private ground truth estimator, we do not known whether we can infer the ability of each user under LDP model and what is the estimation error with respect to the underlying ability of users. Finally, previous work only shows that their methods have better performance than the

[3] Note that in the data mining community this problem is also called Truth Discovery.

private major voting algorithm through experiments on some datasets. However, there is still no theoretically result which shows the priority of their methods formally or mathematically.

In this paper, we partially solve the above theoretical issues. That is, instead of considering the LDP model, in this paper we will focus on one of its relaxations called local attribute differential privacy (LADP) model. This is motivated by the fact that in practice of ground truth inference, instead of keeping the each whole data record of each user private, it is always the case that only a small number of answers given by users may contain sensitive information, which means it is sufficient to protect some attributes of a vector (if we see the set of answers of each user as a vector). LADP corresponds to an adversary cannot infer a single attribute value despite he knows the values of all other attributes and thus is more suitable for ground truth inference. We study the previous issues of ground truth inference in LADP model. In particularly, we propose a method called private Dawid-Skene method which outputs the private truth estimators and private ability of users. Specifically, our contributions can be summarized as the followings.

- We first show that our private Dawid-Skene method is LADP. Then we provide the result on the average error of the private truth estimators w.r.t the ground truth. We show that under some statistical assumptions of the problem and if the initial vector of the algorithm is closed enough to the ground truth, then the average error will be upper bounded by $\exp(-n\tilde{v})$ with high probability, where n is the number of users and $\tilde{v}$ is the term called collective private wisdom which is related to the privacy level ϵ (see Theorem 2 for details).
- We also show that under the same assumptions, the output of private ability of users has the estimation error of $O(\sqrt{\frac{\log m}{m\epsilon^2}})$, where m is the number of tasks with high probability (see Theorem 3 for details).
- Finally, we compare our method with the classical private major voting algorithm. To show the priority of our method, we propose a special instance. We show that the estimation error given by the private major voting error is always greater than the error given by our algorithm, which means the private major voting is always worse than our method on this instance theoretically. See Theorem 4 for details.

2 Preliminaries

In this section, we review the definition of ground truth estimation in crowdsourcing, local differential privacy and the classical Dawid-Skene algorithm.

2.1 Local Differential Privacy

Definition 1 (Differential Privacy [4]). *Given a data universe $\mathcal{X}$, we say that two datasets $D, D' \subseteq \mathcal{X}$ are neighbors if they differ by only one entry, which is denoted as $D \sim D'$. A randomized algorithm $\mathcal{A}$ is (ϵ, δ)-differentially private (DP) if for all neighboring datasets D, D' and for all events S in the output space of $\mathcal{A}$, the following holds*

$$\mathbb{P}(\mathcal{A}(D) \in S) \leq e^{\epsilon}\mathbb{P}(\mathcal{A}(D') \in S) + \delta.$$

When $\delta = 0$, $\mathcal{A}$ is ϵ-differentially private.

Instead of the trusted curator, in Local Differential Privacy model [13], each data provider perturb his/her private data record locally via some differentially private mechanisms before sending it to the curator. It is defined as the follows.

Definition 2. *A randomized algorithm $\mathcal{A}$ is ϵ-locally differentially private (LDP) of for all $x, x' \in \mathcal{X}$ that are different and all events S, the following holds*

$$\mathbb{P}(\mathcal{A}(x) \in S) \leq e^{\epsilon}\mathbb{P}(\mathcal{A}(x') \in S).$$

Note that the LDP can be regarded as a special case of traditional DP where each dataset only contains one tuple. Thus, for the same privacy parameter ϵ, LDP provides a stronger guarantee than DP.

However, in some problems, it is always the case that only some of the attributes are related to users sensitive information. We only need to protect some attributes instead of the whole data record of each user in LDP model. Formally, it is called Local Attribute Differential Privacy (LADP), which is a relaxation of LDP and has been studied in many previous papers, such as [9, 10, 14, 17]. Mathematically it can be defined as the follows.

Definition 3 (Local Attribute Differential Privacy). *A randomized algorithm $\mathcal{A}$ is ϵ-locally is differentially private if for all $x, x' \in \mathcal{X}$ with there is some i where x and x' are differ in the i-th coordinate and all for all events S in the output of $\mathcal{A}$, we have*

$$\mathbb{P}(\mathcal{A}(x) \in S) \leq e^{\epsilon}\mathbb{P}(\mathcal{A}(x') \in S).$$

We note that the only difference between LDP and LADP is in LADP we have an additional restriction on x, x'. LADP corresponds to an adversary cannot infer a single attribute value despite he knows the values of all other attributes.

2.2 Problem Setting

We now start by formally define the problem of **Ground Truth Inference**. Conceptually, there are two parties, sever and user, are involved in the crowdsourced question answering. We assume there are m tasks and n users, each task $j \in [m]$ is independent with other tasks and is associated with a label $y_j^* \in \{0, 1\}$ which is called the **ground truth**. We note that in practice the number of tasks m is much larger than the number of users n, such as the Web and AdultCotent datasets [22]. The users, who represent the individual participants, provide their own answer 0 or 1 to each of these tasks and send them to the server. However, there is one main issue. Due to the quality of the users, these answers are noisy. It is more challenging that the underlying quality of the workers are also unknown. Mathematically, to model the users' quality, [2] proposed the so-called confusion matrix. The confusion matrix for the i-the worker is denoted as

$$\begin{bmatrix} \pi_{00}^{(i)}, \pi_{01}^{(i)} \\ \pi_{10}^{(i)}, \pi_{11}^{(i)} \end{bmatrix}$$

where the number $\pi_{kl}^{(i)}$ represents the probability for the i-th user to give answer l given the ground truth is k. In our paper, we will study a special class of the confusion matrix, where the ability of the i-th user is characterized by the probability of success $p_i^* \in [0,1]$ with the confusion matrix

$$\begin{bmatrix} p_i^*, 1-p_i^* \\ 1-p_i^*, p_i^* \end{bmatrix}.$$

Equivalently, here we will assume that for each user $i \in [n]$, his/her abilities are the same for all the m tasks.

After collecting the users answers, the server aggregate them to derive the final inference and estimation. The goal is not only inferring the truth labels $\{y_j^*\}_{i=1}^m$, but also estimating the abilities of the users, *i.e.,* $\{p_i^*\}_{i=1}^n$.

The main privacy concern of users is that the submitted answers many contain their sensitive information and thus users are not willing to leak these answers to other parties. This prevents users from sharing their own answers with the server. The server, who is assumed to be untrusted, may try to infer additional knowledge of users forms their submitted answers. The unfaithful behavior of server can be driven by financial incentives or other benefits. Motivated by this, it is naturally to study the problem of ground truth inference under LDP model. However, the definition if LDP might be too strong for the problem of ground truth inference. Since in the problem, it is always the case that only some of the tasks are related to users sensitive information. Thus it is sufficient of we can protect these tasks instead of the whole data record of each user in LDP model, which is just the LADP model.

Thus, motivated by the strong need to provide users with privacy protection. In the **Private Ground Truth Inference** problem, we want to design ϵ-LADP algorithms whose outputs $\{y_j\}_{j=1}^n$ and $\{p_i^*\}_{i=1}^n$ are close to $\{y_j^*\}_{i=1}^m$ and $\{p_i^*\}_{i=1}^n$, respectively.

3 Main Method

In this section we will propose our method and analyze its theoretical performance. Before that, we first recall the classical Dawid-Skene method [2].

3.1 Dawid-Skene Method

Now we consider the problem of ground truth inference in the non-private case (see Section 2.2). We first observe that the ability of the works $\{p_i\}_{i=1}^n$ can be easily estimated by using the frequency of success of the workers if the ground truth $\{y_j^*\}_{j=1}^m$ is known. Motivated by this, [2] proposed to estimate $\{p_i\}_{i=1}^n$ by maximizing the marginal likelihood function by giving the ground truth:

$$\begin{aligned}\mathbb{P}(X|y,p) &= \mathit{\Pi}_{j\in[m]}\mathit{\Pi}_{i\in[n]}\mathbb{P}(X_{ij}|y_j,p_i) \\ &= \mathit{\Pi}_{j\in[m]}\mathit{\Pi}_{i\in[n]}p_i^{\mathbb{I}(X_{ij}=y_j)}(1-p_i)^{\mathbb{I}(X_{ij}=1-y_j)}, \qquad (1)\end{aligned}$$

where $\mathbb{I}$ is the indicator function [4]. Integrating out the ground truth with a uniform prior, the marginal likelihood is

$$\mathbb{P}(X|p) = \mathit{\Pi}_{j\in[m]}\big(\frac{1}{2}\mathit{\Pi}_{i\in[n]}p_i^{X_{ij}}(1-p_i)^{1-X_{ij}} \\ + \frac{1}{2}\mathit{\Pi}_{i\in[n]}(1-p_i)^{X_{ij}}p_i^{1-X_{ij}}\big). \quad (2)$$

Thus, the the maximum likelihood estimator (MLE) based on (2) is defined as

$$\hat{p} = \arg\max_p \log P(X|p).$$

After getting the MLE solution $\hat{p} = (\hat{p}_1, \hat{p}_2, \cdots, \hat{p}_n)$, we can plug it into the Bayes formula and get an estimator for the ground truth y^*:

$$\hat{y}_j = \frac{\mathit{\Pi}_{i\in[n]}\hat{p}_i^{X_{ij}}(1-\hat{p}_i)^{1-X_{ij}}}{\mathit{\Pi}_{i\in[n]}\hat{p}_i^{X_{ij}}(1-\hat{p}_i)^{1-X_{ij}} + \mathit{\Pi}_{i\in[n]}(1-\hat{p}_i)^{X_{ij}}\hat{p}_i^{1-X_{ij}}}, \quad (3)$$

Note that we implicitly use the uniform prior in the Bayes formula and the resulting estimator $\hat{y}$ is a soft label, taking value in $[0,1]^m$. Now the pair of estimator $(\hat{p}, \hat{y})$ is the global optimizer of the following objective function.

$$F(p,y) = \sum_i\sum_j y_j\big(X_{ij}\log p_i + (1-X_{ij})\log(1-p_i)\big) + \\ \sum_i\sum_j(1-y_i)\big(X_{ij}\log(1-p_i)+(1-X_{ij})\log p_i) + \sum_j(y_j\log\frac{1}{y_j} + (1-y_j)\log\frac{1}{1-y_j}. \quad (4)$$

[18] showed that optimizing over $\log\mathbb{P}(X|p)$ is equivalent as optimizing over $F(p,y)$, *i.e.,* $(\hat{p}, \hat{y}) = \arg\max F(p,y)$, while the latter one is more tractable. In order to maximize (4), one natural and heuristic way is to iteratively update p an y. That is, given an initial estimator $y^{(0)}$, the t-th step of the iterative algorithm is

$$p^{(t)} = \arg\max F(p, y^{(t-1)}), y^{(t)} = \arg\max F(p^{(t)}, y). \quad (5)$$

Calculating (5) directly, we have the followings:

$$p_i^{(t)} = \frac{1}{m}\sum_{j\in[m]}\big((1-X_{ij})(1-y_j^{(t-1)}) + X_{ij}y_j^{(t-1)}\big), \quad (6)$$

$$y_j^{(t)} \propto \mathit{\Pi}_{i\in[n]}(p_i^{(t)})^{X_{ij}}(1-p_i^{(t)})^{1-X_{ij}}, \quad (7)$$

$$1 - y_j^{(t)} \propto \mathit{\Pi}_{i\in[n]}(p_i^{(t)})^{1-X_{ij}}(1-p_i^{(t)})^{X_{ij}}. \quad (8)$$

Eq. (6)-(8), are given by [2] and are called Dawid-Skene method.

[4] Given an event A, $\mathbb{I}(A) = 1$ if A happens and otherwise it is 0.

3.2 Private Dawid-Skene Estimation

Now we propose the our Private Dawid-Skene method. The idea is that for each user $i \in [n]$ who process answers $(X_{i1}, X_{i2}, \cdots, X_{ij})$, he/she perturbs each answer by the following distribution:

$$\hat{X}_{ij} = \begin{cases} X_{ij} \text{ w.p. } \frac{e^{\epsilon}}{e^{\epsilon}+1} \\ 1 - X_{ij} \text{ w.p. } \frac{1}{e^{\epsilon}+1}. \end{cases} \tag{9}$$

After the server getting these perturbed answers $\{\hat{X}_{ij}\}_{i\in[n],j\in[m]}$, it then performs the Dawid-Skene estimator on these perturbed answers, see Algorithm 1 for details. However, we note that instead of performing (6) for updating the abilities of users, here we perform a projected version, that is

$$p_i^{(t)} = \mathit{\Pi}_{\mathcal{C}(\lambda)} \frac{1}{m} \sum_{j\in[m]} \Big((1 - X_{ij})(1 - y_j^{(t-1)}) + X_{ij} y_j^{t-1} \Big). \tag{10}$$

Where $\mathit{\Pi}_{\mathcal{C}(\lambda)}$ is the projection operator on a interval $\mathcal{C}(\lambda) = [\lambda, 1-\lambda]$ with some small $\lambda > 0$. The motivation is that in the case of the estimator $p_i^{(t)}$ is 0 or 1 for some $i \in [n]$ and $t \in [T]$, $p_i^{(t)}$ will be trapped in its current value, which might be a poor local optimizer. Thus, in order to avoid, we perform the projector operator to keep $p_i^{(t)}$ be slightly away from 0 or 1. Later, we will see that an appropriate value of λ is crucial for the rate of convergence. We note that this operator also has been used and studied in [7].

Also, we note that, after the T-th iteration, instead of releasing the the estimators of the ability $p_i^{(T)}$ directly, we have to post-process them via Step 7 in Algorithm 1. This is due to that, $\{p_i^T\}_{i\in[n]}$ are some biased estimators of the underlying ability $\{p_i^*\}_{i\in[n]}$ since the perturbation procedure in Step 2. Thus, in order to get some useful estimators we need to rescale them. We will see later for the reason of choosing these terms for rescaling.

Finally, since the terms $\{y_j\}_{j\in[m]}$ are soft labels contained in $[0,1]^m$, in order to get hard labels as final answers, we need to do a round procedure in step 8. We will show that it will not effect the error to much.

The following theorem shows that the algorithm is LADP. Not only LADP, it is also easy to see that Algorithm 1 is also $m\epsilon$ locally differentially private.

Theorem 1. *For any given $\epsilon > 0$, Algorithm 1 is ϵ-LADP.*

3.3 Theoretical Guarantees

In this section, we will give the estimation errors of the outputs $\{\hat{p}_i^{(T)}\}_{i\in[n]}$ and $\{y_j^{(T)}\}_{j=1}^m$ to the underlying abilities and ground truth, respectively. Before showing the explicit result, we first introduce some critical quantities.

First, for each user $i \in [n]$, we define the term of **private effective ability** as

$$\hat{\mu}_i = \frac{e^{\epsilon} - 1}{e^{\epsilon} + 1} \mu_i + \frac{1}{e^{\epsilon} + 1}, \tag{11}$$

Algorithm 1 Private Dawid-Skene Method

```
Input: $T$ is the number of iteration, $\epsilon > 0$ is the privacy parameter, $y^{(0)}$ is the initial vector, worker $i \in [n]$ process the answers $X_i = (X_{i1}, \cdots, X_{im}) \in \{0,1\}^m$.
for Each worker $i \in [n]$ do
    Perturb each $X_{ij}, j \in [m]$ by the distribution (9) and get $\hat{X}_{ij}$. Then send $\{\hat{X}_{ij}\}_{j=1}^m$ to the server.
end for
for $t = 1, \cdots, T$ do
    The server perform the updating (10), (7), (8) on $\{\hat{X}_{ij}\}_{i\in[n], j\in[m]}$ and get $\{p_i^{(t)}\}_{i\in[n]}$, $\{y_j^{(t)}\}_{j\in[m]}$.
end for
For each $i \in [n]$, let $\hat{p}_i^{(T)} = \frac{e^\epsilon+1}{e^\epsilon-1}(p_i^{(T)} - \frac{1}{e^\epsilon+1})$.
For each $j \in [m]$, let $\hat{y}_j^{(T)} = \mathbb{I}(y_j^{(T)} \geq \frac{1}{2})$.
return $\hat{p}^{(T)} = \{\hat{p}_i^{(T)}\}_{i\in[n]}$ and $\{\hat{y}_j^{(T)}\}_{j=1}^m$.
```

where μ_i is the **effective ability** proposed by [7]:

$$\mu_i = p_i^* \mathbb{I}\{p_i^* \geq \frac{1}{2}\} + (1 - p_i^*)\mathbb{I}\{p_i^* < \frac{1}{2}\}. \tag{12}$$

Intuitively, μ_i measures how much information we can get from the user i: when $p_i^* > \frac{1}{2}$ it is just the underlying ability, when $p_i < \frac{1}{2}$ then we can use the information to detect and invert the answers. $\hat{\mu}_i$ can be thought as the private version of μ_i due to the effect of perturbation by (9). When the algorithm is extremely private *i.e.,* $\epsilon \to 0$, we can see that $\hat{\mu}_i \to \frac{1}{2}$, that is all the workers become spammers. Equivalently, from (9) we can see that $\mathbb{P}(\hat{X}_{ij} = 0) = \mathbb{P}(\hat{X}_{ij} = 1) = \frac{1}{2}$, which means we cannot get any useful information from the perturbed observations. However, when the algorithm tends to be non-private, that is $\epsilon \to \infty$, we have $\hat{\mu}_i \to \mu_i$, in this case the private effective ability will be the same as the effective ability. We note that for both $\mu_i, \hat{\mu}_i$ are in $[\frac{1}{2}, 1]$.

Now we define the term of **collective private wisdom** $\hat{v}$ as

$$\hat{v} = \frac{1}{n}\sum_{i=1}^{n}(2\hat{\mu}_i - 1)^2. \tag{13}$$

$\hat{v}$ measures the proportion of experts among the crowd under the privacy constraint, when the $\epsilon \to 0$, then $\hat{v} \to 0$ since in the extreme private case we cannot distinguish which one is the expert. When $\epsilon \to \infty$, then $\hat{v} \to v = \frac{1}{n}\sum_i (2\mu_i - 1)^2$, which is the collective wisdom in [7].

Note that the objective function $F(p, y)$ in (4) is non-convex with the fixed $\{\hat{X}_{ij}\}_{i\in[n], j\in[m]}$. Thus, alternating maximization procedures (6) and (7) will only converge to some local minimum. However, in the following theorem, we will show that under the setting of $m \gg n$, with some appropriate initial vector $y^{(0)}$, the iterations $\{y_j^{(t)}\}_{j\in[m]}$ after the first step will be in the neighborhood of the ground truth $\{y_j^*\}_{j\in[m]}$ with high probability.

Theorem 2. *Assume n and m are sufficiently large so that $n \leq m \leq e^n$, $\frac{\log m}{n} \leq \hat{v}$ and the initial vector $y^{(0)}$ satisfies*

$$\frac{1}{m}\sum_{j\in[m]} |y_j^{(0)} - y_j^*| \leq \sqrt{\frac{\log m}{m}}. \tag{14}$$

Whenever the parameter λ of Algorithm 1 are chose in the range

$$\frac{16}{\hat{v}}\sqrt{\frac{\log m}{m}} \leq \lambda \leq \frac{1}{8} - \frac{1}{2}\sqrt{\frac{\log m}{m}}. \tag{15}$$

Then for any $y^* \in \{0,1\}^m$, we have

$$\frac{1}{m}\sum_{j\in[m]} |\hat{y}_j^{(T)} - y_j^*| \leq 2\exp(-\frac{1}{2}n\hat{v}). \tag{16}$$

with probability at least $1 - \frac{C'}{m}$ for some constant $C' > 0$.

From Theorem 2, we can see that as long as the our initial guess has the average error of $\tilde{O}(\frac{1}{\sqrt{m}})$, then for some λ the average error will decreases to $\exp(-\frac{1}{2}n\hat{v})$. We can see that when ϵ deceases, this upper bound will increase, which means the error will be larger. Equivalently, this shows that when the algorithm is more private, the error bound will be larger. When $\epsilon = 0$, the upper bound becomes $\frac{1}{2}$ and will be trivial.

The following theorem states that our algorithm not only can almost infer the ground truth, but also can estimate the users' abilities with some statistical error.

Theorem 3. *Under the assumptions in Theorem 2. For $0 < \epsilon \leq 1$ we have the followings with probability at least $1 - \frac{C'}{m}$ for some $C' > 0$:*

$$\max_{i\in[n]} |\hat{p}_i^{(T)} - p_i^*| \leq 6\sqrt{\frac{\log m}{m\epsilon^2}}, \tag{17}$$

Proof. For (17), due to (23) we can see for each $j \in [n]$

$$|p_j^{(T)} - \hat{p}_j^*| \leq \sqrt{\frac{\log m}{m}} + r^{t-1} \leq 2\sqrt{\frac{\log m}{m}}.$$

By the definition of $\hat{p}_i^*$. We have

$$|p_j^{(T)} - \frac{e^\epsilon - 1}{e^\epsilon + 1}p_j^* + \frac{1}{e^\epsilon + 1}| \leq 2\sqrt{\frac{\log m}{m}}$$

which implies

$$|\hat{p}_j^{(T)} - p_j^*| \leq 2\frac{e^\epsilon + 1}{e^\epsilon - 1}\sqrt{\frac{\log m}{m}} \leq \frac{6}{\epsilon}\sqrt{\frac{\log m}{m}}.$$

Eq. (17) characterize the accuracy for users' abilities from the worst-case. We know the rate of error is $\tilde{O}(\frac{1}{m\epsilon^2})$, which means that it will decreases as the number of tasks increases. Moreover, when the algorithm is more private, the bound will be larger.

4 Comparison with Private Major Voting

In order to show the priority of our method theoretically, in this part, we will compared our algorithm with the most trivial method *i.e.,* private major voting. The algorithm of private major voting is quite simple; the steps of the user side is the same as steps 1-3 in Algorithm 1 while each user send the private answers to the server. After collecting all of the private answers, the server will do major voting and decide the output for each task, that is for all $j \in [m]$

$$\bar{y}_j = \mathbb{I}(\sum_{i\in[n]} \hat{X}_{ij} \geq \frac{n}{2}). \tag{18}$$

Now we will provide a case where the upper bound (16) is lower than the bound of private major voting, which means our algorithm has better performance than private major voting theoretically. Formally, suppose that there are $\lceil n^\delta \rceil$ number of experts, *i.e.,* $p_i^* = 1$ and the left workers are spammers, *i.e.,* $p_i^* = \frac{1}{2}$. Here we assume that $\delta \in (0, \frac{1}{2})$, that is only a small proportion of workers are experts.

Next theorem show that the expected average error of the outputs $\{\bar{y}_j\}_{j\in[m]}$ in (18) of private major voting is larger than the average error in Theorem 2 if ϵ in some range.

Theorem 4. *For any $\epsilon > 0$, private major voting is ϵ-LADP. Moreover, if $\epsilon > \ln \frac{85}{15}$ and $n > C$ for some sufficiently large constant C (only related to δ), then the outputs $\{\bar{y}_j\}_{j\in[m]}$ satisfy*

$$\frac{1}{m}\sum_{j\in[m]} \mathbb{E}|\bar{y}_j - y_j^*| \geq 2\big(\exp(-\frac{1}{2}(\frac{e^\epsilon - 1}{e^\epsilon + 1})^2\lceil n^\delta\rceil)\big) \geq \frac{1}{m}\sum_{j\in[m]} |\hat{y}_j^{(T)} - y_j^*|, \tag{19}$$

where $\{\hat{y}_j^{(T)}\}$ are outputs of Algorithm 1.

5 Related Work

There is much attention on studying crowdsourcing system in LDP model. For example, [20] consider the problem of publishing high dimensional crowdsourced data in LDP model. [11] propose a method which could generating synthetic crowdsourced data via some Privacy-Test. However, their methods are incomparable with ours due to that there utility is different with ours and also there is no theoretical guarantees on their output.

Among all the previous work, maybe [16] and [22] are the most relevant to ours. In [16] the authors propose a two-layer perturbation mechanism based on randomized response to protect users privacy. [22] consider the case where the data is sparse and propose a private mechanism based on the formula of Matrix Factorization and randomized response. However, as we mentioned before, first, their methods can only output private estimators of the ground truth and it is unknown whether they can also estimate the ability of users. Secondly, there is no theoretical guarantees of the average error of the ground truth. Moreover, in all of these work they compared with the private major voting algorithm practically on some datasets and showed that their method have better performance. However, there is no theoretical guarantees on these comparisons. Thus, our work provides some theoretical guarantees which have not been solved in these previous work.

Our method is motivated by the classical Dawid-Skene method [2], which laid a solid foundation in the field of crowdscourcing. Extensions of the framework under a Bayesian setting were investigated by [1]. However, there is no previous study on the private version of Dawid-Skene method. Moreover, compared with the classical Dawid-Skene method, here we need some modifications such as perturbation and projection see Section 3.2 for details.

6 Conclusion

In this paper we study the problem of ground truth inference under local attribute differential privacy model and propose an algorithm called private Dawid-Skene method. Specifically, under some statistical assumptions, we provide the first results on the estimation error of the abilities of the users and the ground truth which have not been studied before. Moreover, we propose an explicit example and show that our method has less error than the private major voting algorithm theoretically, which is the first result on theoretically comparing with private major voting method.

There are still some open problems. For example, the model we consider in the paper is LADP, which is a relaxation of LDP. So the first question is whether we can extend our results to the LDP model. Secondly, as we can see from the paper, all of our theorems needs to assume that the ability for each user is the same for all the tasks. Thus how to extend to the case where these ability are different will be another question. Finally, in our theorems, we need to assume the initial vector is already closed to the underlying parameters. Thus, the problem is that how to find this initial vector privately or how to relax this assumption is still an open problem. We leave these problems as future work.

References

1. Chen, X., Lin, Q., Zhou, D.: Optimistic knowledge gradient policy for optimal budget allocation in crowdsourcing. In: International conference on machine learning. pp. 64–72 (2013)
2. Dawid, A.P., Skene, A.M.: Maximum likelihood estimation of observer error-rates using the em algorithm. Journal of the Royal Statistical Society: Series C (Applied Statistics) **28**(1), 20–28 (1979)
3. Ding, H., Gao, J., Xu, J.: Finding global optimum for truth discovery: Entropy based geometric variance. In: Proc. 32nd International Symposium on Computational Geometry (SoCG 2016) (2016)
4. Dwork, C., McSherry, F., Nissim, K., Smith, A.: Calibrating noise to sensitivity in private data analysis. In: Theory of cryptography conference. pp. 265–284. Springer (2006)
5. Dwork, C., Roth, A., et al.: The algorithmic foundations of differential privacy. Foundations and Trends® in Theoretical Computer Science **9**(3–4), 211–407 (2014)
6. Erlingsson, Ú., Pihur, V., Korolova, A.: Rappor: Randomized aggregatable privacy-preserving ordinal response. In: Proceedings of the 2014 ACM SIGSAC conference on computer and communications security. pp. 1054–1067. ACM (2014)

7. Gao, C., Zhou, D.: Minimax optimal convergence rates for estimating ground truth from crowdsourced labels. arXiv preprint arXiv:1310.5764 (2013)
8. Haney, S., Machanavajjhala, A., Abowd, J.M., Graham, M., Kutzbach, M., Vilhuber, L.: Utility cost of formal privacy for releasing national employer-employee statistics. In: Proceedings of the 2017 ACM International Conference on Management of Data. pp. 1339–1354. ACM (2017)
9. Hardt, M., Roth, A.: Beyond worst-case analysis in private singular vector computation. In: Proceedings of the forty-fifth annual ACM symposium on Theory of computing. pp. 331–340. ACM (2013)
10. Heinze-Deml, C., McWilliams, B., Meinshausen, N.: Preserving privacy between features in distributed estimation. Stat **7**(1), e189 (2018)
11. Huai, M., Wang, D., Miao, C., Xu, J., Aidong, Z.: Privacy-aware synthesizing for crowdsourced data. In: 28th International Joint Conference on Artificial Intelligence (IJCAI 2019) (2019)
12. Huang, Z., Ding, H., Xu, J.: Faster algorithm for truth discovery via range cover. In: Workshop on Algorithms and Data Structures. pp. 461–472. Springer (2017)
13. Kasiviswanathan, S.P., Lee, H.K., Nissim, K., Raskhodnikova, S., Smith, A.: What can we learn privately? SIAM Journal on Computing **40**(3), 793–826 (2011)
14. Kifer, D., Machanavajjhala, A.: No free lunch in data privacy. In: Proceedings of the 2011 ACM SIGMOD International Conference on Management of data. pp. 193–204. ACM (2011)
15. Li, S., Xu, J., Ye, M.: Approximating global optimum for probabilistic truth discovery. In: International Computing and Combinatorics Conference. pp. 96–107. Springer (2018)
16. Li, Y., Miao, C., Su, L., Gao, J., Li, Q., Ding, B., Qin, Z., Ren, K.: An efficient two-layer mechanism for privacy-preserving truth discovery. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 1705–1714. ACM (2018)
17. Lou, J., Cheung, Y.m.: Uplink communication efficient differentially private sparse optimization with feature-wise distributed data. In: Thirty-Second AAAI Conference on Artificial Intelligence (2018)
18. Neal, R.M., Hinton, G.E.: A view of the em algorithm that justifies incremental, sparse, and other variants. In: Learning in graphical models, pp. 355–368. Springer (1998)
19. O'Neill, L., Dexter, F., Zhang, N.: The risks to patient privacy from publishing data from clinical anesthesia studies. Anesthesia & Analgesia **122**(6), 2017–2027 (2016)
20. Ren, X., Yu, C.M., Yu, W., Yang, S., Yang, X., McCann, J.A., Philip, S.Y.: $\{LoPub\}$: High-dimensional crowdsourced data publication with local differential privacy. IEEE Transactions on Information Forensics and Security **13**(9), 2151–2166 (2018)
21. Shevtsova, I.: On the absolute constants in the berry-esseen type inequalities for identically distributed summands. arXiv preprint arXiv:1111.6554 (2011)
22. Sun, H., Dong, B., Wang, H.W., Yu, T., Qin, Z.: Truth inference on sparse crowdsourcing data with local differential privacy. In: 2018 IEEE International Conference on Big Data (Big Data). pp. 488–497. IEEE (2018)
23. Tang, J., Korolova, A., Bai, X., Wang, X., Wang, X.: Privacy loss in apple's implementation of differential privacy on macos 10.12. CoRR **abs/1709.02753** (2017)
24. Zhang, Y., Chen, X., Zhou, D., Jordan, M.I.: Spectral methods meet em: A provably optimal algorithm for crowdsourcing. The Journal of Machine Learning Research **17**(1), 3537–3580 (2016)
25. Zheng, Y., Li, G., Li, Y., Shan, C., Cheng, R.: Truth inference in crowdsourcing: Is the problem solved? Proceedings of the VLDB Endowment **10**(5), 541–552 (2017)
26. Zhi, S., Yang, F., Zhu, Z., Li, Q., Wang, Z., Han, J.: Dynamic truth discovery on numerical data. In: 2018 IEEE International Conference on Data Mining (ICDM). pp. 817–826. IEEE (2018)

A Omitted Proofs

Proof (Proof of Theorem 1). Now consider $X_i, X_i' \in \{0,1\}^m$ differ in the j-th coordinate, *i.e.,* $X_i = (X_{i1}, X_{i2}, \cdots, X_{ij}, \cdots, X_{im})$ and $X_i = (X_{i1}, X_{i2}, \cdots, X'_{ij}, \cdots, X_{im})$. For any $S \in \{0,1\}^m$, by the independence and the definition of (9) we have

$$\frac{\mathbb{P}(\hat{X}_i \in S)}{\mathbb{P}(\hat{X}_i' \in S)} = \frac{\mathbb{P}(\hat{X}_{ij} = S_j)}{\mathbb{P}(\hat{X}'_{ij} = S_j)} \tag{20}$$

when $X_{ij} = 1$ and $X_{ij} = 0$ and $S_j = 1$ then (20) equals e^ϵ. When S_j the then (20) equals $\frac{1}{e^\epsilon} \leq e^\epsilon$. The same for the case where $X_{ij} = 0$ and $X_{ij} = 1$. Thus in total we can see that (20) less equals than e^ϵ, which satisfies the definition of LADP. Moreover, due to the post-processing property of differential privacy [5], we know that Algorithm 1 is LADP.

Proof (Proof of Theorem 2). By the definition of $\hat{X}_{ij}$ and the assumption, we can represent it as

$$\hat{X}_{ij} = y_j^* T_{ij} + (1 - y_j^*)(1 - T_{ij}) \tag{21}$$

where T_{ij} is a Bernoulli random variable with parameter

$$\hat{p}_i^* = \frac{e^\epsilon}{e^\epsilon + 1} p_i^* + \frac{1}{e^\epsilon + 1}(1 - p_i^*) = \frac{e^\epsilon - 1}{e^\epsilon + 1} p_i^* + \frac{1}{e^\epsilon + 1}. \tag{22}$$

We notice that T_{ij} means that the i-th worker answers the j-th task correctly.

We also define the projected version of $\hat{p}_j^*$ as

$$\hat{p}^*_{\lambda,i} = \lambda \mathbb{I}(\hat{p}_i^* < \lambda) + \hat{p}_i^* \mathbb{I}(\lambda \leq \hat{p}_i^* \leq 1 - \lambda) + (1 - \lambda)\mathbb{I}(\hat{p}_i^* > 1 - \lambda).$$

To proof Theorem 2, we first proof a stronger claim that for each iteration $t \geq 1$, $\{y_j^{(t)}\}_{j \in [m]}$ satisfies Eq. (16) with probability at least $1 - \frac{C'}{m}$.

We denote the error of $\{y_j^{(t)}\}_{j \in [m]}$ as r^t, that is

$$r^t = \frac{1}{m} \sum_{j \in [m]} |y_j^{(t)} - y_j^*|.$$

By assumption (14) we know $r^0 \leq \sqrt{\frac{\log m}{m}}$. We first prove the following lemma:

Lemma 1. *Define the events*

$$E_1 = \{\max_{i \in [n]} |\frac{1}{m} \sum_{j \in [m]} (T_{ij} - \hat{p}_i^*)| \leq \sqrt{\frac{\log m}{m}}\}.$$

$$E_2 = \{\max_{j \in [m]} |\sum_{i \in [n]} (T_{ij} - \hat{p}_i^*) \log \frac{\hat{p}^*_{\lambda,i}}{1 - \hat{p}^*_{\lambda,i}}| \leq 2 \log(\frac{1}{\lambda}) \sqrt{n \log m}\}.$$

Then $\mathbb{P}(E_1 \bigcap E_2) \geq 1 - \frac{C'}{m}$ *for some* $C' > 0$.

Proof (Proof of Lemma 1). To proof this, we recall the Hoeffiding's inequality

Lemma 2 (Hoeffiding's inequality). *For independent bounded random variables $\{X_i\}_{i\in[n]}$ satisfying $X_i \in [a_i, b_i]$ for all $i \in [n]$, we have for any $t \geq 0$*

$$\mathbb{P}\big(|\frac{1}{n}\sum_{i\in[n]}(X_i - \mathbb{E}X_i)| > t\big) \leq 2\exp(\frac{-2n^2t^2}{\sum_{i\in[n]}(b_i - a_i)^2}).$$

Note that for the Event E_1, by Lemma 2, we have $\mathbb{P}(E_1) \geq 1 - \frac{C_1}{m}$ for some $C_1 > 0$.

For the event E_2, we note that by the definition of $\hat{p}^*_{\lambda,i}$ we have

$$\log\frac{\hat{p}^*_{\lambda,i}}{1-\hat{p}^*_{\lambda,i}} \leq \log\frac{1-\lambda}{\lambda} \leq \log\frac{1}{\lambda}.$$

Thus, by Lemma 2, we know there is a $C_2 > 0$, where $\mathbb{P}(E_2) \geq 1 - \frac{C_2}{m}$.

In the following we will always assume events E_1 and E_2 in Lemma 1 hold. Next we will prove the following lemma:

Lemma 3. *Under the event E_1, as long as $2\lambda + r^{t-1} \leq \frac{1}{4}$ and $m \geq 9$, we have for all $t \geq 1$:*

$$\max_{i\in[n]}|\log\frac{p_i^{(t)}}{1-p_i^{(t)}} - \log\frac{\hat{p}^*_{\lambda,i}}{1-\hat{p}^*_{\lambda,i}}| \leq \frac{2}{\lambda}\sqrt{\frac{\log m}{m}} + \frac{2}{\lambda}r^{t-1}.$$

Proof (Proof of Lemma 3). We note that from Eq. (6) and Eq. (10) on $\{\hat{X}_{ij}\}_{i\in[n],j\in[m]}$ we can get

$$p_i^{(t)} = \lambda\mathbb{I}(\bar{p}_i^{(t)} < \lambda) + \bar{p}_i^{(t)}\mathbb{I}(\lambda \leq \bar{p}_i^{(t)} \leq 1-\lambda) + (1-\lambda)\mathbb{I}(\bar{p}_i^{(t)} > 1-\lambda).$$

Where $\bar{p}_i^{(t)}$ is the value of (6), *i.e.,* the vector before projecting. By the definitions (21) and (6) we can get the following via simple calculations:

$$|\bar{p}_i^{(t)} - \hat{p}_i^*| \leq |\frac{1}{m}\sum_j(T_{ij} - \hat{p}_i^*)| + r^{t-1}. \tag{23}$$

To show (23), by definition of $\bar{p}_i^{(t)}$ we have

$$p_i^{(t)} = \frac{1}{m}\sum_{j\in[m]}\big((1-\hat{X}_{ij})(1-y_j^{(t-1)}) + \hat{X}_{ij}y_j^{(t-1)}\big). \tag{24}$$

Now we fix $j \in [m]$ and assume that $y_j^* = 1$, then by (21) we have $\hat{X}_{ij} = T_{ij}$, we can get

$$|(1-\hat{X}_{ij})(1-y_j^{(t-1)}) + \hat{X}_{ij}y_j^{(t-1)} - \hat{p}_i^*| \quad = |2T_{ij}y_j^{(t-1)} - T_{ij} - y_j^{(t-1)} + 1 - p_i^*|. \tag{25}$$

When $T_{ij} = 0$, (25) is $|y_j^{(t-1)} - 1 + p_i^*| \leq |p_i^*| + |y_j^{(t-1)} - y_j^*|$. When $T_{ij} = 1$, (25) is $|y_j^{(t-1)} - p_i^*| \leq |1 - p_i^*| + ||y_j^{(t-1)} - y_j^*|$. Thus in total we have (25) less than $|T_{ij} - p_i^*| + ||y_j^{(t-1)} - y_j^*|$. The same for the case when $y_j^* = 0$.

Taking the average from 1 to m we can get (23).

Now we have for each $i \in [n]$

$$|\log \frac{p_i^{(t)}}{1 - p_i^{(t)}} - \log \frac{\hat{p}^*_{\lambda,i}}{1 - \hat{p}^*_{\lambda,i}}| \leq \frac{\frac{p_i^{(t)}}{1-p_i^{(t)}} - \frac{\hat{p}^*_{\lambda,i}}{1-\hat{p}^*_{\lambda,i}}}{\min\{\frac{p_i^{(t)}}{1-p_i^{(t)}}, \frac{\hat{p}^*_{\lambda,i}}{1-\hat{p}^*_{\lambda,i}}\}} \tag{26}$$

$$\leq \frac{2}{\lambda}|p_i^{(t)} - \hat{p}^*_{\lambda,i}| \tag{27}$$

$$\leq \frac{2}{\lambda}|\bar{p}_i^{(t)} - \hat{p}_i^*| + \frac{4}{\lambda}\mathbb{I}(|\bar{p}_i^{(t)} - \hat{p}_i^*| > 1 - 2\lambda) \tag{28}$$

$$\leq \frac{2}{\lambda}(|\frac{1}{m}\sum_j (T_{ij} - \hat{p}_i^*)| + r^{t-1}) + \frac{4}{\lambda}\mathbb{I}(|\frac{1}{m}\sum_j (T_{ij} - \hat{p}_i^*)| > \frac{3}{4}) \tag{29}$$

$$\leq \frac{2}{\lambda}\sqrt{\frac{\log m}{m}} + \frac{2}{\lambda}r^{t-1}. \tag{30}$$

Where the first inequality (26) is due to the following inequality

$$|\log x - \log y| \leq \frac{|x - y|}{\min\{x, y\}}.$$

The inequality (27) is due to that $\lambda \leq p_i^{(t)}, \hat{p}^*_{\lambda,i} \leq 1 - \lambda$ and simple calculation.

The inequality (28) is due to the following. When $|\bar{p}_i^{(t)} - \hat{p}_i^*| > 1 - 2\lambda$, then $|p_i^{(t)} - \hat{p}^*_{\lambda,i}| \leq 2$. Otherwise by the definition we have either $\bar{p}_i^{(t)}$ or $\hat{p}_i^*$ is in the interval $[\lambda, 1 - \lambda]$, thus we have $|p_i^{(t)} - \hat{p}^*_{\lambda,i}| \leq |\bar{p}_i^{(t)} - \hat{p}_i^*|$ due to the property of contraction of the projection. The inequality (29) is due to the following. By (23) we have

$$\begin{aligned}
&\mathbb{I}(|\bar{p}_i^{(t)} - \hat{p}_i^*| > 1 - 2\lambda) \\
&\leq \mathbb{I}(|\frac{1}{m}\sum_j (T_{ij} - \hat{p}_i^*)| + r^{t-1} > 1 - 2\lambda) \\
&= \mathbb{I}(|\frac{1}{m}\sum_j (T_{ij} - \hat{p}_i^*)| > 1 - 2\lambda - r^{t-1}) \\
&\leq \mathbb{I}(|\frac{1}{m}\sum_j (T_{ij} - \hat{p}_i^*)| > \frac{3}{4}), \qquad (31)
\end{aligned}$$

where the last inequality is due to the assumption that $2\lambda + r^{t-1} \leq \frac{1}{4}$.

The inequality (29) is due to the assumption of the event E_1 in Lemma 1 holds.

Thus, we get the proof.

Now we back to the proof of Theorem 2.

Since we already know that $r^0 \leq \sqrt{\frac{\log m}{m}}$ and we want to show it hold for all r^t.

We will prove it by induction, assume $r^{t-1} \leq \sqrt{\frac{\log m}{m}}$ holds. Denote the terms $A_j^t, B_j^t, j \in [m]$ as

$$
\begin{aligned}
A_j^t &= \log \mathit{\Pi}_{i\in[n]} (p_i^{(t)})^{\hat{X}_{ij}} (1-p_i^{(t)})^{1-\hat{X}_{ij}} = \sum_i \left(\hat{X}_{ij}\log p_i^{(t)} + (1-\hat{X}_{ij})\log(1-p_i^{(t)})\right) \\
B_j^t &= \log \mathit{\Pi}_{i\in[n]} (1-p_i^{(t)})^{\hat{X}_{ij}} (p_i^{(t)})^{1-\hat{X}_{ij}} = \sum_i \left(\hat{X}_{ij}\log(1-p_i^{(t)}) + (1-\hat{X}_{ij})\log p_i^{(t)}\right).
\end{aligned}
$$

Then by the definition of $\{y_j^{(t)}\}_{j=1}^m$ we have

$$
\begin{aligned}
r^t &= \frac{1}{m}\sum_j |y_j^{(t)} - y_j^*| = \frac{1}{m}\sum_j |\frac{\exp(A_j^t)}{\exp(A_j^t) + \exp(B_j^t)} - y_j^*| \\
&= \frac{1}{m}\sum_j |\frac{\exp(A_j^t - B_j^t)}{\exp(A_j^t - B_j^t) + 1} - y_j^*| \\
&= \frac{1}{m}\sum_j \frac{1}{1 + \exp(\sum_i (2T_{ij}-1)\log\frac{p_i^{(t)}}{1-p_i^{(t)}})} \\
&\leq \frac{1}{m}\sum_j \exp(-\sum_i (2T_{ij}-1)\log\frac{p_i^{(t)}}{1-p_i^{(t)}}) \\
&\leq \frac{1}{m}\sum_j \exp(-\sum_i (2T_{ij}-1)\log\frac{\hat{p}^*_{\lambda,i}}{1-\hat{p}^*_{\lambda,i}} + \frac{4n}{\lambda}\sqrt{\frac{\log m}{m}}) \\
&\leq \frac{1}{m}\sum_j \exp(-\sum_i (2\hat{p}^*_i - 1)\log\frac{\hat{p}^*_{\lambda,i}}{1-\hat{p}^*_{\lambda,i}}) \exp(\frac{4n}{\lambda}\sqrt{\frac{\log m}{m}} + 4\log\frac{1}{\lambda}\sqrt{n\log m}).
\end{aligned}
$$

Where the equalities are followed by the direct computation. The second inequality is by Lemma 3 and the assumption of $r^{t-1} \leq \sqrt{\frac{\log m}{m}}$, the third inequality is due to Lemma 1.

For the exponent in the first term we have

$$\begin{aligned}
&\sum_i (2\hat{p}_i^* - 1)\log\frac{\hat{p}_{\lambda,i}^*}{1-\hat{p}_{\lambda,i}^*}\\
&= (\sum_{i:\hat{p}_i^*<\lambda} + \sum_{i:\lambda\le\hat{p}_i^*\le 1-\lambda} + \sum_{i:\hat{p}_i^*>1-\lambda})(2\hat{p}_i^* - 1)\log\frac{\hat{p}_{\lambda,i}^*}{1-\hat{p}_{\lambda,i}^*}\\
&\ge \big(|\{i:\hat{p}_i^*<\lambda\}| + |\{i:\hat{p}_i^*>1-\lambda\}|\big)(1-2\lambda)\log\frac{1-\lambda}{\lambda}\\
&+ \sum_{i:\lambda\le\hat{p}_i^*\le 1-\lambda}(2\hat{p}_i^* - 1)\log\frac{\hat{p}_{\lambda,i}^*}{1-\hat{p}_{\lambda,i}^*}\\
&\ge \big(|\{i:\hat{p}_i^*<\lambda\}| + |\{i:\hat{p}_i^*>1-\lambda\}|\big) + \sum_{i:\lambda\le\hat{p}_i^*\le 1-\lambda}(2\hat{p}_i^* - 1)^2\\
&\ge \sum_i (2\hat{p}_i^* - 1)^2
\end{aligned}$$

In the following we will show that $(2\hat{p}_i^* - 1)^2 = (2\hat{\mu}_i - 1)^2$, by this we have $\sum_i (2\hat{p}_i^* - 1)^2 = n\hat{v}$. This is due to the following equation:

$$\hat{\mu}_i = \hat{p}_i^*\mathbb{I}(\hat{p}_i^* \ge \frac{1}{2}) + (1-\hat{p}_i^*)\mathbb{I}(\hat{p}_i^* < \frac{1}{2}). \tag{32}$$

Thus, in total we have

$$\begin{aligned}
r^t &\le \exp(\frac{4n}{\lambda}\sqrt{\frac{\log m}{m}} + 4\log\frac{1}{\lambda}\sqrt{n\log m} - n\hat{v}) &(33)\\
&\le \exp(-\frac{1}{2}n\hat{v}) \le \sqrt{\frac{\log m}{m}}. &(34)
\end{aligned}$$

Where the second inequality is due to the assumption on the range of λ.

Next, due to the rounding procedure (Step 8 of Algorithm 1) and

$$\begin{aligned}
|\mathbb{I}(y_j \ge \frac{1}{2}) - 1| &\le 2|y_j - 1|\\
|\mathbb{I}(y_j < \frac{1}{2}) - 1| &\le 2|y_j - 0|
\end{aligned}$$

We have $\frac{1}{m}\sum_{j\in[m]}|\hat{y}_j^{(T)} - y_j^*| \le 2\exp(-\frac{1}{2}n\hat{v})$.

Proof (Proof of Theorem 4). To proof Theorem 4, we need the Berry-Essen Lemma in [21]:

Lemma 4. *Let $X_1, \cdots, X_n$ be i.i.d random variables with mean 0 and variance σ^2. Define the function $F_n(t) = \mathbb{P}(\frac{1}{\sigma\sqrt{n}}\sum_i X_i \le t)$. Then we have*

$$\sup_{t\in\mathbb{R}}|F_n(t) - \Phi(t)| \le \frac{c\mathbb{E}|X_1|^3}{\sigma\sqrt{n}}, \tag{35}$$

where $c < 0.4748$ and $\Phi(t)$ is the cumulative distribution function of the standard Gaussian distribution $\mathcal{N}(0,1)$.

By the definition of majority voting and the definition in (21) we have $|\bar{y}_j - y_j^*| = \mathbb{I}(\frac{1}{n}\sum_i T_{ij} < \frac{1}{2})$. To prove this, we first consider the case where $y_j^* = 1$. Then by (21) we have

$$\begin{aligned} |\bar{y}_j - y_j^*| &= |\mathbb{I}(\sum_{i\in[n]} \hat{X}_{ij} \geq \frac{n}{2}) - 1| \\ &= |\mathbb{I}(\sum_{i\in[n]} T_{ij} \geq \frac{n}{2}) - 1| = \mathbb{I}(\frac{1}{n}\sum_i T_{ij} < \frac{1}{2}). \end{aligned}$$

The same for the case where $y_j^* - 1$.

Thus

$$\begin{aligned} \frac{1}{m}\sum_{j\in[m]} \mathbb{E}|\bar{y}_j - y_j^*| &= \frac{1}{m}\sum_{j\in[m]} \mathbb{P}(\frac{1}{n}\sum_i T_{ij} < \frac{1}{2}) \\ &= \mathbb{P}(\frac{1}{n}\sum_i T_i < \frac{1}{2}), \end{aligned}$$

where $\{T_i\}_{i\in[n]}$ are independent Bernoulli random variable with parameter $\hat{p}_i^*$ in (22).

By the definition (22), we known that if $p_i^* = \frac{1}{2}$ then $\hat{p}_i^* = \frac{1}{2}$, if $p_i^* = 1$ then $\hat{p}_i^* = \frac{e^\epsilon}{e^\epsilon+1}$. W.l.o.g we assume $p_i^* = \frac{1}{2}$ for $i \leq n - \lceil n^\delta \rceil$. By Lemma 4 with $\mathbb{E}[T_i - \frac{1}{2}] = 0$, $\text{Var}(T_i - \frac{1}{2}) = \frac{1}{4}$ and $\mathbb{E}|T_i - \frac{1}{2}|^3 = \frac{1}{8}$ for $i \leq n - \lceil n^\delta \rceil$ we have

$$\sup_t |\mathbb{P}\{\frac{2}{\sqrt{n - \lceil n^\delta \rceil}} \sum_{i \leq n - \lceil n^\delta \rceil} (T_i - \frac{1}{2}) \leq t\} - \Phi(t)| \leq \frac{(n - \lceil n^\delta \rceil)^{-\frac{1}{2}}}{16}. \tag{36}$$

Also by direct calculation we have

$$\begin{aligned} \mathbb{P}(\frac{1}{n}\sum_i T_i < \frac{1}{2}) &\geq \mathbb{P}\{\frac{2}{\sqrt{n - \lceil n^\delta \rceil}} \sum_{i \leq n - \lceil n^\delta \rceil} (T_i - \frac{1}{2}) \\ &\geq -\frac{\lceil n^\delta \rceil}{\sqrt{n - \lceil n^\delta \rceil}}\} \times \mathbb{P}\{T_i = 1, \forall i > n - \lceil n^\delta \rceil\} \end{aligned}$$

Thus by (36) we have

$$\mathbb{P}(\frac{1}{n}\sum_i T_i < \frac{1}{2}) \geq (\frac{e^\epsilon}{e^\epsilon + 1})^{\lceil n^\delta \rceil} \times \{\Phi(-\frac{\lceil n^\delta \rceil}{\sqrt{n - \lceil n^\delta \rceil}}) - \frac{(n - \lceil n^\delta \rceil)^{-\frac{1}{2}}}{16}\}. \tag{37}$$

We know that since $\delta < \frac{1}{2}$, thus for sufficiently large n we have $\Phi(-\frac{\lceil n^\delta \rceil}{\sqrt{n - \lceil n^\delta \rceil}}) \geq \frac{1}{4}$ and $\frac{(n - \lceil n^\delta \rceil)^{-\frac{1}{2}}}{16} \leq \frac{1}{8}$, which are due to that

$$\begin{aligned} &\lim_{n\to\infty} \Phi(-\frac{\lceil n^\delta \rceil}{\sqrt{n - \lceil n^\delta \rceil}}) = \Phi(0) = \frac{1}{2}, \\ &\lim_{n\to\infty} \frac{(n - \lceil n^\delta \rceil)^{-\frac{1}{2}}}{16} = 0. \end{aligned}$$

Thus

$$\mathbb{P}(\frac{1}{n}\sum_{i} T_i < \frac{1}{2}) \geq (\frac{e^{\epsilon}}{e^{\epsilon}+1})^{\lceil n^{\delta}\rceil}\frac{1}{8}.$$

On the other side by Theorem 2 we have

$$\frac{1}{m}\sum_{j\in[m]} |\hat{y}_j^{(T)} - y_j^*| \leq 2\big(\exp(-\frac{1}{2}(\frac{e^{\epsilon}-1}{e^{\epsilon}+1})^2\lceil n^{\delta}\rceil)\big).$$

Now we will show that for large enough n:

$$2\big(\exp(-\frac{1}{2}(\frac{e^{\epsilon}-1}{e^{\epsilon}+1})^2\lceil n^{\delta}\rceil)\big) \leq (\frac{e^{\epsilon}}{e^{\epsilon}+1})^{\lceil n^{\delta}\rceil}\frac{1}{8}. \tag{38}$$

Denote $v = \frac{e^{\epsilon}}{e^{\epsilon}+1} \in [0.85, 1)$, it is equivalent to show

$$\frac{\exp(-\frac{1}{2}(2v-1)^2\lceil n^{\delta}\rceil)}{v^{\lceil n^{\delta}\rceil}} \leq \frac{1}{16} \tag{39}$$

Thus, it is sufficient if we can show the following

$$\lim_{n\to\infty}\frac{\exp(-\frac{1}{2}(2v-1)^2\lceil n^{\delta}\rceil)}{v^{\lceil n^{\delta}\rceil}} = 0. \tag{40}$$

we note LHS of (40) equals to $\exp((-\frac{1}{2}(2v-1)^2 - \log v)\lceil n^{\delta}\rceil)$, we will show $f(v) = \frac{1}{2}(2v-1)^2 + \log v > 0$ under our assumption on ϵ. This is due to that $f(v)$ is an increasing function, it is easy to see that $f(0.85) > 0$. Thus we proof Eq. (40).